心理科学进展　2019, Vol. 27, No. 7, 1205–1214
Advances in Psychological Science　DOI: 10.3724/SP.J.1042.2019.01205

·研究前沿(Regular Articles)·

面孔表情和声音情绪信息整合加工的脑机制*

李　萍　张明明　李帅霞　张火垠　罗文波
(辽宁师范大学脑与认知神经科学研究中心，大连　116029)

摘　要　在现实生活中，有效的情绪识别往往依赖于不同通道间的信息整合(如，面孔、声音)。本文梳理相关研究认为，面孔表情和声音情绪信息在早期知觉阶段即产生交互作用，且初级感知觉皮层负责两者信息的编码；而在晚期决策阶段，杏仁核、颞叶等高级脑区完成对情绪信息内容的认知评估整合；此外，神经振荡活动在多个频段上的功能耦合促进了跨通道情绪信息整合。未来研究需要进一步探究两者整合是否与情绪冲突有关，以及不一致的情绪信息在整合中是否有优势，探明不同频段的神经振荡如何促进面孔表情和声音情绪信息整合，以便更深入地了解面孔表情和声音情绪信息整合的神经动力学基础。

关键词　情绪；整合效应；ERP；神经振荡；fMRI
分类号　B842

1　引言

在日常交流中，个体需要有效识别对话者情绪表达的变化，以便快速做出反应(Van Kleef, 2009)。情绪信息通过多个感觉通道进行表达，特别是面孔表情和声音(Campanella & Belin, 2007; Klasen, Chen, & Mathiak, 2012)。因此，个体需要同时提取双通道的情绪信息，并将它们整合起来，在面对面的互动中形成统一而连贯的感知对象(de Gelder & Vroomen, 2000)。研究发现，大脑能自动地完成这一过程(Chen, Han, Pan, Luo, & Wang, 2016; Chen, Pan, et al., 2016; Chen, Pan, Wang, Zhang, & Yuan, 2015)，从而优化信息处理，减少不确定性(Doi & Shinohara, 2015; Kuhn, Wydell, Lavan, McGettigan, & Garrido, 2017; Noy et al., 2017)。

有关双通道情绪性信息，尤其是面孔表情和声音情绪信息整合的研究，主要从以下两方面进行考察：一方面，比较双通道情绪性信息在一致和不一致的条件下，是否存在情绪信息的一致性促进作用(Klasen, Kenworthy, Mathiak, Kircher, & Mathiak, 2011; Mileva, Tompkinson, Watt, & Burton, 2018; Muller, Cieslik, Turetsky, & Eickhoff, 2012; Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000)；另一方面，比较单通道与双通道的情绪信息处理效率(Jessen & Kotz, 2011; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007)。早期行为结果表明，面孔表情和声音情绪信息存在整合加工现象。具体表现为，在双通道条件下，个体对情绪信息识别的反应更快，准确率更高，具有促进作用(张亮，孙向红，张侃, 2009; de Gelder & Vroomen, 2000; Klasen et al., 2012; Schelenz et al., 2013)。此外，一致的面孔表情和情绪性声音能够促进情绪识别(Campanella & Belin, 2007; Klasen, Kreifelts, Chen, Seubert, & Mathiak, 2014)。近年来，大量研究借助事件相关电位(event-related potential, ERP)、功能性磁共振成像(functional magnetic resonance imaging, fMRI)等认知神经科学技术考察面孔表情和声音情绪信息整合加工的认知过程和神经机制。并且提出双通道信息整合应当满足这些标准：首先，相对于单通道感知情绪信息，双通道情绪信息同时加工时具有一定的优势；其次，具有超加性，面孔表情和声音的双通道情绪信息同时加工的效应大于任一单通道，甚至大于所有单通道效应之和，即使是面孔信息较为模糊时，

收稿日期：2018-05-17
* 国家自然科学基金项目(31871106)资助。
通信作者：罗文波, E-mail: luowb@lnnu.edu.cn

也可整合成为一致情绪客体，如 AV > [unimodal auditory (A) + unimodal visual (V)]，或理解为 1 + 1 > 2；再次，遵循颠倒效应规则(the principal of inverse effectiveness)，参与整合的单通道传递信息较为模糊，或是双通道信息都较为模糊时，整合效应更为显著(Stein & Stanford, 2008; Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009; 王苹，潘治辉，张立洁，陈煦海, 2015)。

本文根据实验技术手段(ERP、EEG、MEG、fMRI)对认知神经科学研究的结果进行梳理，以揭示面孔表情和声音情绪信息整合加工的脑机制，并进一步提出未来研究中值得关注的问题。

2 面孔表情和声音情绪信息整合的电生理学研究

电生理测量如脑电图(ERP、EEG)和脑磁图(MEG)等技术由于其高时间分辨率的优势，有助于解决面孔表情和声音情绪信息整合加工的时间进程这一问题。本文将根据时间进程，逐步阐述这一整合加工的变化过程。

2.1 面孔表情和声音情绪信息整合的 ERP 研究

面孔表情和声音情绪信息的整合在 100 ms 左右就已经开始，情绪一致的面孔表情和声音比不一致的情绪刺激诱发了波幅更大的 N100 (Pourtois et al., 2000)。大量研究同时发现，面孔表情和声音情绪性信息在类似的时间窗口出现了交互效应(Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014; Romero, Senkowski, & Keil, 2015; van Wassenhove, Grant, & Poeppel, 2005)，表明面孔表情和声音的情绪信息在早期就进行了知觉编码。此外，与单通道呈现面孔或声音刺激相比，双通道的面孔表情和声音情绪诱发了波幅更小的 N1 (Jessen & Kotz, 2011; Kokinous, Kotz, Tavano, & Schroger, 2015)。进一步研究发现，即使以更复杂的呈现方式(声音、面部表情和身体姿势)，双通道的情绪信息均比单通道条件下诱发更小的 N1 波幅(Yeh, Geangu, & Reid, 2016)。这可能表明不同通道的情绪信息存在相互作用，整合发生在多个加工层次，即在感官、认知和反应选择层面。总之，面孔表情和声音情绪信息整合在 100 ms 左右就已经开始。

亦有电生理研究表明，P2 波幅的变化是双通道中情绪信息冗余(redundant)、一致(coherent)、整合(convergent)处理的相关指标(Balconi & Carrera, 2011; Ho, Schroger, & Kotz, 2015)。P2 出现在刺激后约 200 ms，不一致的面孔表情和声音情绪刺激诱发更小的 P2 波幅(Pourtois, Debatisse, Despland, & de Gelder, 2002)。具体表现为，当愤怒面孔与中性声音同时呈现时，P2 波幅减小；且中性面孔与愤怒的声音相结合使 P2 的波幅最小(Balconi & Carrera, 2011; Ho et al., 2015)。另有 Doi 和 Shinohara (2015)的研究表明，通过连续的闪光抑制方式，将高兴、恐惧和中性的面孔在阈下呈现给被试，同时播放快乐、恐惧的声音，要求被试根据声音线索完成情绪识别任务。结果发现，枕颞区的 P2 受面孔和声音的交互调节。在高兴面孔–笑声组合和中性面孔–笑声组合中，P2 的波幅没有显著差异，而恐惧面孔和笑声组合诱发 P2 波幅显著小于中性面孔与笑声组合刺激。进一步表明在情绪不一致情况时，阈下呈现的恐惧面孔也能促进对笑声的加工。但是也有研究发现，仅在情绪一致的条件下，面孔表情才能促进对声音情绪的识别(Collignon et al., 2008; Focker, Gondan, & Roder, 2011; Zinchenko, Obermeier, Kanske, Schroger, & Kotz, 2017)，所以有必要进一步探明在双通道条件下，不一致情绪刺激对情绪识别是否也具有促进作用。

Paulmann, Jessen 和 Kotz (2009)采用身份识别任务，直接对比单通道和视听结合情绪刺激，发现面孔表情和声音情绪刺激诱发更小 P2 波幅。另有研究发现，情绪强度也调节双通道情绪信息整合，在低强度条件下，愤怒面孔表情和声音情绪一致的刺激诱发的 P2 波幅显著大于单通道愤怒面孔或声音刺激(Pan, Liu, Luo, & Chen, 2017)。P2 虽然是对情绪刺激进行快速检测的重要指标，但其在面孔表情和声音情绪信息整合加工过程中的作用更要复杂一些(Balconi & Carrera, 2011; Ho et al., 2015; Huang et al., 2012; Klasen et al., 2011)。综上可见，被试在同时注意面孔表情和声音情绪信息的条件下，面孔表情可以影响大脑对声音情绪信息的加工，即使在忽略面孔刺激(在阈下或视觉皮层受损的患者中呈现)时，也能影响声音情绪信息的知觉。这表明，在面孔表情和声音信息呈现 200 ms 内，是对情绪刺激低水平的视觉特征进行整合加工。

有关面孔表情和声音情绪信息整合的研究发

现，当情绪刺激呈现 400 ms，额叶皮层区域会出现一个特异性负波——N400，这与情绪信息的分类、评估等精细加工有关(Campanella et al., 2010; Liu, Rigoulot, & Pell, 2015; Paulmann et al., 2009; Proverbio & De Benedetto, 2018)。其反映了在晚期时间窗口内，面孔表情和声音情绪信息整合发生了质的变化(Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000; Hernandez-Gutierrez et al., 2018; Olofsson & Polich, 2007; Schupp et al., 2007)。Paulmann 和 Pell (2010a)运用启动范式，发现不一致面孔表情和声音情绪刺激可诱发典型的 N400，且 N400 波幅受刺激呈现时间长短的调节。他们的另一项研究采用情绪 Stroop 任务，比较不同文化群体在面孔表情和声音情绪整合时的异同，要求被试在情绪一致或者不一致的条件下，忽略另一种通道信息，仅判断其中一个通道的情绪类型。结果表明，情绪不一致的面孔表情和声音准确率较低，诱发更大的 N400 波幅，也出现群体性差异(Liu et al., 2015)。以上研究中，N400 不但可以区分情绪与中性刺激，而且能辨别不同类型的情绪，体现了晚期加工阶段大脑对情绪信息的精细加工(Proverbio & De Benedetto, 2018; Strelnikov, Foxton, Marx, & Barone, 2015)。

综上所述，面孔表情和声音情绪信息在刺激呈现的 200 ms 内已经进行整合，但整合时间随实验任务不同而延伸，在 200 ms 后很长一段时间内都能发生整合。且在神经电生理层面，以更严格的整合(超加性效应和颠倒效应规则)标准，发现双通道情绪信息整合为一致情绪客体的证据。

2.2 面孔表情和声音情绪信息整合的 EEG 和 MEG 研究

除了事件相关的脑电研究外，一些研究采用时频分析来研究面孔表情和声音情绪信息整合中的神经振荡模式。从生态效度来看，人际沟通中的情感表达是随着时间推移不断发生变化的。因此，支持情绪整合的神经系统，必须能灵活地适应并加工来自多通道快速变化的情绪信息。大量研究一致表明，神经振荡活动在这一过程中扮演着重要角色(Gao et al., 2013; Kokinous et al., 2015; Simon & Wallace, 2018; Yang & Lin, 2017)。下文将梳理面孔表情和声音情绪信息整合的神经振荡的相位特征以及脑区间神经振荡的相干特征(是对大脑两个区域同时记录到的EEG信号中相同节律成分进行相位稳定性测量)，探明其是如何在神经振荡层面进行整合的。

Theta 振荡调动大脑和躯体资源来评估和应对情绪变化(Chen et al., 2015; Chen, Yang, Gan, & Yang, 2012)，在面孔表情和声音情绪信息整合过程中起关键的作用，且常与情绪面孔记忆有关。直接对比单通道和双通道呈现面孔表情和声音情绪信息的研究发现，双通道诱发了更大的 Theta 振荡同步化(Chen et al., 2010; Hagan, Woods, Johnson, Green, & Young, 2013)。其中 Chen, Pan 等(2016)的研究表明，Theta 振荡与双通道情绪信息整合加工有关的证据。实验者通过交叉剪辑的方式将面孔表情和声音从中性到愤怒配对变化随机呈现给被试，有 4 个实验条件：1)面孔表情变化，2)声音情绪变化，3)面孔和声音情绪都变化，4)都不变化。被试任务是忽略声音刺激，仅是判断面孔情绪。结果表明，所有类型的情绪变化都会促进 Theta 同步振荡，即被试可以有效的整合来自两个通道的信息，大脑对面孔情绪的判断受到同时呈现的声音情绪的影响。在所有的条件下，与面孔和声音单通道效应之和相比，双通道同时加工情绪信息诱发更大的 Theta 同步振荡，表现出了超加性效应。该效应与识别情绪变化的准确性所观察到的双通道的超加性效应呈正相关。这表明在检测情绪变化过程中，面孔表情和声音情绪信息整合可能是由 Theta 振荡所介导的。值得注意的是 Theta 振荡在高级认知皮层情绪信息整合中作用，研究者给被试同时呈现恐惧和中性的面部表情与声音。结果发现，面孔表情和声音情绪一致的刺激相对不一致的刺激引发颞上沟(superior temporal sulcus, STS)脑区的 Theta 振荡增强，说明双通道冗余情绪信息引发 Theta 频带同步化增强，Theta 振荡同步化受高级认知皮层调节的情绪增益控制(emotional gain control)，情绪一致信息对整合加工的促进作用(Symons, El-Deredy, Schwartze, & Kotz, 2016)。

面孔表情和声音情绪信息整合出现在早期感知皮层还是高级认知皮层，Alpha 振荡支持哪一种观点呢？Jessen 和 Kotz (2011)的研究表明，与单通道的信息相比，情绪一致的面孔表情和声音信息在刺激呈现后 300~700 ms 间，对额叶 Alpha 振荡有较强的抑制作用；随后在一项情绪识别任务中，实验条件分别是：视听(auditory-visual, AV)、

听觉(A)、视觉(V)，采集获得脑磁图数据。为考察单通道和双通道交叉整合，分别确定了 AV > A 和 AV > V 两种对比。在 AV > A 对比中，观察到早期 Alpha 振荡仅限于枕部，而颞叶皮层没有改变；但在 AV > V 的对比中，显示只有在颞叶皮层 Alpha 振荡增强，而视觉皮层没有变化。因此，在感知联合皮层没有发现面孔表情和声音情绪信息整合效应。相反，AV > A 和 AV > V 对比中额叶 Alpha 振荡都显著增强支持了面孔表情和声音情绪信息在高级认知皮层中整合的观点(Chen et al., 2010)。

但是，Schelenz 等人(2013)的研究基于同步脑电−功能磁共振成像(EEG-fMRI)技术，将独立脑电成分的时间过程与血流动力学响应联系起来，视听刺激是动态愤怒、高兴和中性面孔/声音，包括情绪一致和不一致的条件。结果发现，在两种条件下，在面孔表情和声音情绪刺激呈现后 200~400 ms 和 400~700 ms 期间都使枕区 Alpha 振荡同步化；但在早期(200~400 ms)的时间窗口内，仅在情绪一致的条件下，发现前额叶皮层的 Alpha 振荡同步化增强。其可能的原因是，在早期知觉自动化加工阶段，双通道的冗余信息加工占用较少的认知资源；也可能是基于自上而下或刺激驱动(stimulus-driven)的注意控制，对有意注意的选择过滤机制的依赖较少，从而表现出显著的促进作用。这种促进效应不仅局限于前额区，也有可能来自额叶内侧和外侧评估网络(Klasen et al., 2012)。另有对 Alpha 振荡同步激活的研究表明视觉和听觉情绪信息的整合是大规模的动态网络，不是简单的单一视觉网络和听觉网络相加(Ding, Li, Wang, & Luo, 2017; Fingelkurts, Fingelkurts, & Seppo, 2005; Jia, Peng, Li, Hua, & Zhao, 2012; Symons et al., 2016)。因此，Alpha 振荡是一种间接测量皮层活动的方法，在刺激呈现后 200~400 ms 和 400~700 ms 期间，枕叶 Alpha 振荡同步化很可能表明在早期感知觉皮层和晚期高级认知皮层上都对情绪信号进行有效处理，感知皮层对其进行感知处理，高级认知皮层进行消除情绪刺激歧义等精细加工，这表明面孔表情和声音情绪信息整合最主要的脑机制可能是参与神经元间的时间同步化过程。

另有 Gamma 振荡与面孔表情和声音情绪信息的多层级整合加工密切相关(Gao et al., 2013; Tallon-Baudry & Bertrand, 1999)。已有研究发现，与单通道信息相比，双通道情绪信息整合引起听觉区和 STS 脑区 Gamma 频段相干的增加，即新皮质运用脑区间的交流来促进双通道信息整合(Maier, Chandrasekaran, & Ghazanfar, 2008)。使用 MEG 的研究发现，与情绪不一致的刺激相比情绪一致的刺激诱发更强 Gamma 振荡。首先在 200~700 ms 的时间窗口内在右侧颞上回(right superior temporal gyrus, STG)激活，并于 250~750 ms 在左侧颞中回(left middle temporal gyrus, MTG)观察到最高峰值，最后激活扩散至左侧颞上回(left STG)，对多个区域之间同时记录到的 MEG 信号中相同节律成分进行相位稳定性测量发现存在正相关(Hagan et al., 2013)。这一过程中表明，来自双通道的情绪信息会发生时间性耦合，这种耦合会引发额叶调控皮层产生相位相关的高频同步化放电。也就说 Gamma 同步化振荡实现了暂时性的功能连接，从而促进跨通道的信息整合(Tang, Wu, & Shen, 2016)。进而，Kumar, Kumar, Roy 和 Banerjee (2018)的研究提出，额叶皮层区 Gamma 振荡功能连接增强是面孔表情和声音情绪信息整合的一个标志。

总之，面孔表情和声音情绪信息的整合和多频段神经振荡有关。且面孔表情和声音一致的情绪信息在神经振荡指标上表现出显著的超加性和颠倒效应(Chen, Pan, et al., 2016; Hagan et al., 2009; Jessen & Kotz, 2011)。在面孔表情和声音整合加工的研究中，Theta 振荡能区分情绪和中性面孔并且可以编码情绪刺激。相比于单通道情绪信息加工，双通道情绪刺激同时加工时，Theta 波段在 STS 脑区显著增强。Alpha 振荡主要是在视觉成分的研究中发现，且 Alpha 同步振荡网络比其他网络更密集，是因为面孔表情被用作视觉信息处理的运动特征，说明视觉信息是视听整合的关键(Jia et al., 2012; Symons et al., 2016)。在诸如杏仁核、颞上沟、颞中回、初级视觉皮层、丘脑等情绪处理区域中观察到 Gamma 同步振荡(Jia et al., 2012; Jochen, Ingo, Hermann, Klaus, & Werner, 2005; Lin, Liu, Liu, & Gao, 2015)。这表明多个皮层 Gamma 同步化振荡，从而促进跨通道的信息整合，可能与初级视觉皮层的 Gamma 振荡受到了颞叶皮层的调控有关。进一步说明，多个脑区间 Gamma 振荡的相干是面孔表情和声音情绪信息整合时脑区间信息交流的核心模式。但是，关于面孔表情和

声音情绪信息整合在神经振荡层面如何整合成为一致的情绪客体，还未有统一认识，缺乏系统的研究，今后需要对这一重要问题进行深入探讨。

3 面孔表情和声音情绪信息整合的 fMRI 研究

利用较高空间分辨率的 fMRI 技术，探明面孔表情和声音情绪信息在哪里整合的问题。进一步了解面孔表情和声音情绪信息整合所涉及的高级认知联合皮层以及高级认知联合皮层和感知皮层之间的交互作用方式，且通过比较单通道与多通道刺激的激活脑区发现多通道刺激的加工区域并非是各个单通道刺激加工区域的简单叠加(张亮 等, 2009)。

来自神经影像的研究，为颞叶在面孔表情和声音情绪信息整合加工中进行精细编码提供了支持(Belyk, Brown, Lim, & Kotz, 2017; Delle-Vigne, Kornreich, Verbanck, & Campanella, 2015; Kreifelts et al., 2007; Kreifelts, Ethofer, Shiozawa, Grodd, & Wildgruber, 2009; Zhu & Beauchamp, 2017)。其中，STS 在面孔表情和声音情绪信息整合中扮演着重要角色。同单通道刺激相比，面孔表情和声音情绪刺激更容易引起后颞上沟(posterior STS)的激活(Robins, Hunyadi, & Schultz, 2009)。随后研究将 STS 的功能进行了细化。例如，STS 的主干部分(the trunk section)是最显著的声音特异区，STS 后端上升分支(the posterior terminal ascending branch)是面孔特异区，STS 的分叉处(the bifurcation)则是面孔表情和声音整合的特异区(Kreifelts et al., 2009)。值得注意的是，与单通道情绪刺激相比，双通道的情绪刺激激活左侧后颞上沟、双侧后颞上回、STS、MTG，其中 pSTG (posterior STG)的激活最为显著(Kreifelts, Ethofer, Huberle, Grodd, & Wildgruber, 2010)。并且，当被试完成情绪类别判断任务时，恐惧面孔表情和声音情绪一致的刺激显著激活左侧 pSTS (Ethofer, Anders, et al., 2006)，说明不同的情绪有其不同的加工脑区。但是，当面孔表情和声音情绪信息一致或者不一致时都有 STS 激活，这也表明不同情绪类型的面孔表情和声音刺激可能共享同一整合神经网络，面孔表情和声音情绪信息发生自动化整合(Hagan et al., 2013; Symons et al., 2016)。

这些研究表明，颞叶很可能存在一个“神经元群体” (neuronal population)，是由分散的单个感觉神经元和多感觉神经元组成，而多感觉神经元可能在情感处理的背景下整合信息。此外，情感信号的视听整合在前颞上沟(anterior STS)达到高峰，与面部和声音特异区重叠，这意味着在面孔表情和声音感知加工过程中，潜在的声音和面部特异的神经元群体之间可能存在相互作用(Kreifelts et al., 2009)。但右侧 pSTS 并不完全是由多感知神经元组成，面孔表情和声音情绪信息整合效应部分是由这些神经元介导的。所以，上述研究表明颞叶皮层是面孔表情和声音情绪信息进行整合的重要脑区，并可能通过反馈回路影响到感觉联合皮层及初级感觉皮层的信息整合。

Dolan, Morris 和 de Gelder (2001)首次发现，杏仁核是面孔表情和声音情绪信息整合的关键脑区。在其研究中，主试在播放快乐或恐惧声音时，同时呈现快乐或恐惧面孔，要求被试忽略声音刺激，仅识别面孔表情。fMRI 的结果显示，恐惧面孔表情和声音刺激更多激活了杏仁核(Park et al., 2010)。Epperson, Amin, Ruparel, Gur 和 Loughead (2012) 的研究也出现了相同的结果。这说明不管个体注意的重点如何，面孔表情和声音情绪信息可以自动整合(Doi & Shinohara, 2015)。此外，大量研究发现，与单通道面孔刺激相比，恐惧的面孔表情和声音情绪刺激显著激活了右侧杏仁核，快乐的面孔表情和声音整合激活左侧杏仁，说明了不同情绪类型的面孔表情和声音有不同的加工脑区(Armony & Dolan, 2000; Klasen et al., 2012; Pourtois, Thut, de Peralta, Michel, & Vuilleumier, 2005)。最后，相比情绪性面孔表情和声音整合，中性的面孔表情和声音整合活动在左杏仁核中显著减弱，这表明左侧杏仁核进行了情绪分类、评估等精细加工(Calvo, Beltran, & Fernandez-Martin, 2014)。

与单通道面孔和声音刺激相比，双通道面孔表情和声音情绪刺激诱发了更强的丘脑活动(Kreifelts et al., 2007)。随后的研究发现，与不一致的面孔表情和声音情绪信息相比，一致的刺激导致丘脑的显著激活(Klasen et al., 2011)。这一结果与丘脑在处理情绪性相关信息的作用是一致的(Kober et al., 2008)。当被试完成情绪分类任务时，与恐惧面孔表情和声音刺激相比，快乐面孔声音信息激活左侧丘脑(Klasen et al., 2011)。因此，说明

在感觉联合皮层区域(如，丘脑)能对面孔表情和声音情绪刺激进行编码，并对的情绪特征进行整合。

总之，已有研究确定了多个脑功能区域，包括颞叶、杏仁核、丘脑、梭状回、脑岛等是面孔表情和声音情绪信息整合中心(Klasen et al., 2012; Symons et al., 2016; Yaple, Vakhrushev, & Jolij, 2016)。面孔表情和声音情绪信息整合不是在单一大脑皮层进行，而是存在多脑区的相互作用。这可能是因为面孔和声音情绪信息整合需要更多的大脑皮层参与，需要对其表达的复杂信息进行精细编码，以适应社会交往和人际互动的需求。

4 总结以及展望

综上所述，关于面孔表情和声音情绪信息整合，在时间进程、相关脑区以及如何整合为一致的情绪客体等方面已取得诸多共识。即面孔表情和声音情绪信息在视听整合过程的多个阶段、多个层级和多个脑区相互作用。神经震荡信号变化遵循整合的超加性和颠倒效应原则，这也为判断面孔表情和声音情绪信息整合成一致情绪客体提供新的依据。首先，在早期自动化加工阶段，听觉和视觉感受器接收到面孔表情和声音情绪信息，并把它们传递到皮层下组织区域，然后向前把情绪信息传递给初级感觉皮层，同时进一步向前传递到感觉联合皮层区域(如，丘脑)。在此阶段，N1 与 P2 反映了对面孔表情和声音情绪信息在呈现 200 ms 内能自动发生整合，是对情绪刺激低水平的视觉特征进行加工；同时，来自双通道的情绪信息常会发生时间性耦合，这种耦合会引发关联性感觉皮层脑区和接受信号的感觉联合区产生相位相关的高频同步化放电，感觉联合皮层的神经振荡同步化，进而使情绪信息产生交互作用。情绪信息在早期感知觉皮层激活之后，最后在晚期时间窗口内，进行评估“情绪意义”等精细加工，高级认知皮层区域(如，额叶)对所有进入的情绪信息开始初步整合，并可能通过反馈机制影响到感知联合皮层及初级感觉皮层，实现进一步的跨区域信息整合，面孔表情和声音情绪信息整合发生了质的变化。

同时，未来研究可以关注以下几个方面：

首先，明确面孔表情和声音情绪性信息的整合效应是否与情绪一致性存在必然关联。以往研究表明，与不一致的情感信息相比，只注意同一种情绪，面孔表情和声音中的冗余信息会导致更快的情感分类(Calvo & Nummenmaa, 2016; Collignon et al., 2008; Focker et al., 2011; Klasen et al., 2012; Kumar et al., 2018; Muller et al., 2012)。但是也有研究表明不一致的情绪刺激也能促进面孔表情和声音的整合加工。例如，Ethofer, Pourtois 和 Wildgruber (2006)要求被试忽略声音信息，仅对面孔情绪评分，结果发现伴随着恐惧声音的情况下，更易把中性面孔的判定为恐惧面孔；Doi 和 Shinohara (2015)的研究在呈现恐惧的面孔表情情况下，对笑声的识别更快更准确。这些研究结果是相冲突的。基于这些发现，面孔表情和声音情绪信息整合是否与情感冲突有关？如果有关，是否意味着面孔表情和声音情绪性信息的整合过程是多个独立过程的组合，其中一些过程容易受到情绪一致的信息调节，而另一些则可能受到不一致的情绪信息的影响呢？这些问题都需要进一步研究。

其次，深入探究神经振荡在面孔表情和声音情绪信息中的作用。虽然已有研究发现同时呈现恐惧或厌恶的面孔表情和声音情绪信息时，在 STS 区出现 Theta 和 Gamma 振荡的超加性效应，在加工中性面孔表情和声音刺激时，不会出现该效应(Hagan et al., 2009; Hagan et al., 2013)。但是 Chen, Pan 等人(2016)研究显示，与仅呈现面孔或声音情绪信息相比，双通道的情绪信息仅诱发 Theta 波段的超加性效应并非 Gamma 波段。且目前研究还未发现 Alpha 振荡的这种超加性效应。所以，关于神经振荡在面孔表情和声音情绪整合中的作用还不明确，仍需要进一步系统探明面孔表情和声音情绪信息的神经振荡的相位特征以及脑区间神经振荡的相干特征。

参考文献

张亮，孙向红，张侃. (2009). 情绪信息的多通道整合. *心理科学进展, 17*(6), 1133–1138.

王苹，潘治辉，张立洁，陈煦海. (2015). 动态面孔和语音情绪信息的整合加工及神经生理机制. *心理科学进展, 23*(7), 1109–1117.

Armony, J. L., & Dolan, R. J. (2000). Modulation of attention by threat stimuli: An fMRI study. *Journal of Cognitive Neuroscience*, 53–53.

Balconi, M., & Carrera, A. (2011). Cross-modal integration of emotional face and voice in congruous and incongruous

pairs: The P2 ERP effect. *Journal of Cognitive Psychology, 23*(1), 132–139.

Belyk, M., Brown, S., Lim, J., & Kotz, S. A. (2017). Convergence of semantics and emotional expression within the IFG pars orbitalis. *Neuroimage, 156*, 240–248.

Calvo, M. G., Beltran, D., & Fernandez-Martin, A. (2014). Processing of facial expressions in peripheral vision: Neurophysiological evidence. *Biological Psychology, 100*, 60–70.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion, 30*(6), 1081–1106.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11*(12), 535–543.

Campanella, S., Bruyer, R., Froidbise, S., Rossignol, M., Joassin, F., Kornreich, C., ... Verbanck, P. (2010). Is two better than one? A cross-modal oddball paradigm reveals greater sensitivity of the P300 to emotional face-voice associations. *Clinical Neurophysiology, 121*(11), 1855–1862.

Chen, X. H., Edgar, J. C., Holroyd, T., Dammers, J., Thoennessen, H., Roberts, T. P. L., & Mathiak, K. (2010). Neuromagnetic oscillations to emotional faces and prosody. *European Journal of Neuroscience, 31*(10), 1818–1827.

Chen, X. H., Han, L. Z., Pan, Z. H., Luo, Y. M., & Wang, P. (2016). Influence of attention on bimodal integration during emotional change decoding: ERP evidence. *International Journal of Psychophysiology, 106*, 14–20.

Chen, X. H., Pan, Z. H., Wang, P., Yang, X. H., Liu, P., You, X. Q., & Yuan, J. J. (2016). The integration of facial and vocal cues during emotional change perception: EEG markers. *Social Cognitive and Affective Neuroscience, 11*(7), 1152–1161.

Chen, X. H., Pan, Z. H., Wang, P., Zhang, L. J., & Yuan, J. J. (2015). EEG oscillations reflect task effects for the change detection in vocal emotion. *Cognitive Neurodynamics, 9*(3), 351–358.

Chen, X. H., Yang, J. F., Gan, S. Z., & Yang, Y. F. (2012). The contribution of sound intensity in vocal emotion perception: Behavioral and electrophysiological evidence. *PLoS One, 7*(1), e30278.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research, 1242*, 126–135.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology, 52*(2), 95–111.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion, 14*(3), 289–311.

Delle-Vigne, D., Kornreich, C., Verbanck, P., & Campanella, S. (2015). The P300 component wave reveals differences in subclinical anxious-depressive states during bimodal oddball tasks: An effect of stimulus congruence. *Clinical Neurophysiology, 126*(11), 2108–2123.

Ding, R., Li, P., Wang, W., & Luo, W. (2017). Emotion processing by ERP combined with development and plasticity. *Neural Plasticity, 2017*(2), 5282670.

Doi, H., & Shinohara, K. (2015). Unconscious presentation of fearful face modulates electrophysiological responses to emotional prosody. *Cerebral Cortex, 25*(3), 817–832.

Dolan, R. J., Morris, J. S., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences of the United States of America, 98*(17), 10006–10010.

Epperson, C. N., Amin, Z., Ruparel, K., Gur, R., & Loughead, J. (2012). Interactive effects of estrogen and serotonin on brain activation during working memory and affective processing in menopausal women. *Psychoneuroendocrinology, 37*(3), 372–382.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., ... Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *Neuroimage, 30*(2), 580–597.

Ethofer, T., Pourtois, G., & Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Progress in Brain Research, 156*(6), 345–361.

Fingelkurts, A. A., Fingelkurts, A. A., & Seppo, K. H. N. (2005). Functional connectivity in the brain--Is it an elusive concept? *Neuroscience & Biobehavioral Reviews, 28*(8), 827-836.

Focker, J., Gondan, M., & Roder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta Psychologica, 137*(1), 36–47.

Gao, Z. F., Goldstein, A., Harpaz, Y., Hansel, M., Zion-Golumbic, E., & Bentin, S. (2013). A magnetoencephalographic study of face processing: M170, gamma-band oscillations and source localization. *Human Brain Mapping, 34*(8), 1783–1795.

Hagan, C. C., Woods, W., Johnson, S., Calder, A. J., Green, G. G. R., & Young, A. W. (2009). MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus. *Proceedings of the National Academy of Sciences of the United States of America, 106*(47), 20010–20015.

Hagan, C. C., Woods, W., Johnson, S., Green, G. G. R., &

Young, A. W. (2013). Involvement of right STS in audio-visual integration for affective speech demonstrated using MEG. *PLoS One, 8*(8), e70648.

Hernandez-Gutierrez, D., Abdel Rahman, R., Martin-Loeches, M., Munoz, F., Schacht, A., & Sommer, W. (2018). Does dynamic information about the speaker's face contribute to semantic speech processing? ERP evidence. *Cortex, 104*, 12–25.

Ho, H. T., Schroger, E., & Kotz, S. A. (2015). Selective attention modulates early human evoked potentials during emotional face-voice processing. *Journal of Cognitive Neuroscience, 27*(4), 798–818.

Huang, X. Q., Zhang, J., Liu, J., Sun, L., Zhao, H. Y., Lu, Y. G., ... Li, J. (2012). C-reactive protein promotes adhesion of monocytes to endothelial cells via NADPH oxidase-mediated oxidative stress. *Journal of Cellular Biochemistry, 113*(3), 857–867.

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *Neuroimage, 58*(2), 665–674.

Jia, G., Peng, X., Li, Y., Hua, S., & Zhao, X. J. (2012). The oscillatory activities and its synchronization in auditory-visual integration as revealed by event-related potentials to bimodal stimuli. *Proceedings of SPIE - The International Society for Optical Engineering, 8291*(1), 52.

Jochen, K., Ingo, H., Hermann, A., Klaus, M., & Werner, L. (2005). Hearing lips: Gamma-band activity during audiovisual speech perception. *Cerebral Cortex, 15*(5), 646-653.

Klasen, M., Chen, Y. H., & Mathiak, K. (2012). Multisensory emotions: Perception, combination and underlying neural processes. *Reviews in the Neurosciences, 23*(4), 381–392.

Klasen, M., Kenworthy, C. A., Mathiak, K. A., Kircher, T. T. J., & Mathiak, K. (2011). Supramodal representation of emotions. *Journal of Neuroscience, 31*(38), 15218–15218.

Klasen, M., Kreifelts, B., Chen, Y. H., Seubert, J., & Mathiak, K. (2014). Neural processing of emotion in multimodal settings. *Frontiers in Human Neuroscience, 8*(8), 822.

Knowland, V. C. P., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. C. (2014). Audio-visual speech perception: A developmental ERP investigation. *Developmental Science, 17*(1), 110–124.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage, 42*(2), 998–1031.

Kokinous, J., Kotz, S. A., Tavano, A., & Schroger, E. (2015). The role of emotion in dynamic audiovisual integration of faces and voices. *Social Cognitive and Affective Neuroscience, 10*(5), 713–720.

Kokinous, J., Tavano, A., Kotz, S. A., & Schroeger, E. (2017). Perceptual integration of faces and voices depends on the interaction of emotional content and spatial frequency. *Biological Psychology, 123*, 155–165.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *Neuroimage, 37*(4), 1445–1456.

Kreifelts, B., Ethofer, T., Huberle, E., Grodd, W., & Wildgruber, D. (2010). Association of trait emotional intelligence and individual fMRI-activation patterns during the perception of social signals from voice and face. *Human Brain Mapping, 31*(7), 979–991.

Kreifelts, B., Ethofer, T., Shiozawa, T., Grodd, W., & Wildgruber, D. (2009). Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice-and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia, 47*(14), 3059–3066.

Kuhn, L. K., Wydell, T., Lavan, N., McGettigan, C., & Garrido, L. (2017). Similar representations of emotions across faces and voices. *Emotion, 17*(6), 912–937.

Kumar, G. V., Kumar, N., Roy, D., & Banerjee, A. (2018). Segregation and integration of cortical information processing underlying cross-modal perception. *Multisensory Research, 31*(5), 481–500.

Lin, Y. F., Liu, B. L., Liu, Z. W., & Gao, X. R. (2015). EEG gamma-band activity during audiovisual speech comprehension in different noise environments. *Cognitive Neurodynamics, 9*(4), 389–398.

Liu, P., Rigoulot, S., & Pell, M. D. (2015). Culture modulates the brain response to human expressions of emotion: Electrophysiological evidence. *Neuropsychologia, 67*, 1–13.

Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. A. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology, 18*(13), 963–968.

Mileva, M., Tompkinson, J., Watt, D., & Burton, A. M. (2018). Audiovisual integration in social evaluation. *Journal of Experimental Psychology: Human Perception and Performance, 44*(1), 128–138.

Muller, V. I., Cieslik, E. C., Turetsky, B. I., & Eickhoff, S. B. (2012). Crossmodal interactions in audiovisual emotion processing. *Neuroimage, 60*(1), 553–561.

Noy, D., Mouta, S., Lamas, J., Basso, D., Silva, C., & Santos, J. A. (2017). Audiovisual integration increases the intentional step synchronization of side-by-side walkers. *Human Movement Science, 56*, 71–87.

Olofsson, J. K., & Polich, J. (2007). Affective visual event-related potentials: Arousal, repetition, and

time-on-task. *Biological Psychology, 75*(1), 101-108.

Pan, Z. H., Liu, X., Luo, Y. M., & Chen, X. H. (2017). Emotional intensity modulates the integration of bimodal angry expressions: ERP evidence. *Frontiers in Neuroscience, 11*.

Park, J. Y., Gu, B. M., Kang, D. H., Shin, Y. W., Choi, C. H., Lee, J. M., & Kwon, J. S. (2010). Integration of cross-modal emotional information in the human brain: An fMRI study. *Cortex, 46*(2), 161–169.

Paulmann, S., Jessen, S., & Kotz, S. A. (2009). Investigating the multimodal nature of human communication insights from ERPs. *Journal of Psychophysiology, 23*(2), 63–76.

Paulmann, S., & Pell, M. D. (2010a). Contextual influences of emotional speech prosody on face processing: How much is enough? *Cognitive Affective & Behavioral Neuroscience, 10*(2), 230–242.

Paulmann, S., & Pell, M. D. (2010b). Dynamic emotion processing in Parkinson's disease as a function of channel availability. *Journal of Clinical and Experimental Neuropsychology, 32*(8), 822–835.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport, 11*(6), 1329–1333.

Pourtois, G., Debatisse, D., Despland, P. A., & de Gelder, B. (2002). Facial expressions modulate the time course of long latency auditory brain potentials. *Cognitive Brain Research, 14*(1), 99–105.

Pourtois, G., Thut, G., de Peralta, R. G., Michel, C., & Vuilleumier, P. (2005). Two electrophysiological stages of spatial orienting towards fearful faces: Early temporo-parietal activation preceding gain control in extrastriate visual cortex. *Neuroimage, 26*(1), 149–163.

Proverbio, A. M., & De Benedetto, F. (2018). Auditory enhancement of visual memory encoding is driven by emotional content of the auditory material and mediated by superior frontal cortex. *Biological Psychology, 132*, 164–175.

Robins, D. L., Hunyadi, E., & Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain and Cognition, 69*(2), 269–278.

Romero, Y. R., Senkowski, D., & Keil, J. (2015). Early and late beta-band power reflect audiovisual perception in the McGurk illusion. *Journal of Neurophysiology, 113*(7), 2342–2350.

Schelenz, P. D., Klasen, M., Reese, B., Regenbogen, C., Wolf, D., Kato, Y., & Mathiak, K. (2013). Multisensory integration of dynamic emotional faces and voices: Method for simultaneous EEG-fMRI measurements. *Frontiers in Human Neuroscience, 7*(1), 729.

Schupp, H. T., Stockburger, J., Codispoti, M., Junghoefer, M., Weike, A. I., & Hamm, A. O. (2007). Selective visual attention to emotion. *Journal of Neuroscience, 27*(5), 1082–1089.

Simon, D. M., & Wallace, M. T. (2018). Integration and temporal processing of asynchronous audiovisual speech. *Journal of Cognitive Neuroscience, 30*(3), 319–337.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9*(4), 255–266.

Stein, B. E., Stanford, T. R., Ramachandran, R., Jr Perrault, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: Alternative criteria, models, and inverse effectiveness. *Experimental Brain Research, 198*(2-3), 113–126.

Strelnikov, K., Foxton, J., Marx, M., & Barone, P. (2015). Brain prediction of auditory emphasis by facialS expressions during audiovisual continuous speech. *Brain Topography, 28*(3), 494–505.

Symons, A. E., El-Deredy, W., Schwartze, M., & Kotz, S. A. (2016). The functional role ofneural oscillations in non-verbal emotional communication. *Frontiers in Human Neuroscience, 10*, 239.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences, 3*(4), 151–162.

Tang, X. Y., Wu, J. L., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience and Biobehavioral Reviews, 61*, 208–224.

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science, 18*(3), 184–188.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181–1186.

Yang, C. Y., & Lin, C. P. (2017). Magnetoencephalography study of different relationships among low- and high-frequency-band neural activities during the induction of peaceful and fearful audiovisual modalities among males and females. *Journal of Neuroscience Research, 95*(1-2), 176–188.

Yaple, Z. A., Vakhrushev, R., & Jolij, J. (2016). Investigating emotional top down modulation of ambiguous faces by single pulse TMS on early visual cortices *Frontiers in Neuroscience, 10*, 305.

Yeh, P. w., Geangu, E., & Reid, V. (2016). Coherent emotional perception from body expressions and the voice. *Neuropsychologia, 91*, 99–108.

Zhu, L. L., & Beauchamp, M. S. (2017). Mouth and voice: A relationship between visual and auditory preference in the human superior temporal sulcus. *Journal of Neuroscience, 37*(10), 2697–2708.

Zinchenko, A., Obermeier, C., Kanske, P., Schroger, E., & Kotz, S. A. (2017). Positive emotion impedes emotional but not cognitive conflict processing. *Cognitive Affective & Behavioral Neuroscience, 17*(3), 665–677.

The integration of facial expression and vocal emotion and its brain mechanism

LI Ping; ZHANG Mingming; LI Shuaixia; ZHANG Huoyin; LUO Wenbo

(*Research Center of Brain and Cognitive Neuroscience, Liaoning Normal University, Dalian* 116029, *China*)

Abstract: The integration of various emotional information from different modalities (e.g., face and voice) plays an important role in our interpersonal communication. In order to understand its brain mechanism, more and more researchers found that the interaction between facial expression and vocal emotional information begins in the early stage of perception, and the integration of emotional information content occurs in the late decision-making stage. In the early stage, the primary sensory cortex is responsible for encoding information; while in the late stage, the amygdala, temporal lobe and other advanced brain regions are responsible for cognitive evaluation. In addition, the functional coupling of oscillation activities on multiple frequency bands facilitates the integration of emotional information cross channels. Future research needs to explore whether facial expression and vocal emotional information integration is associated with emotional conflict, and whether inconsistent emotional information has advantages. Lastly, we should find out how the neural oscillations of different frequency bands promotes the integration of facial expression and vocal emotional information, so as to further understand its dynamic basis.

Key words: emotion; integration effect; ERP; neural oscillation; fMRI